Weekly Report

1 Done

1.1 Paper Revision

Keep going. It is planned to be done before August.

1.2 Learning and Reading

Recent research advances on interactive machine learning

https://arxiv.org/pdf/1811.04548.pdf

Supporting Handoff in Asynchronous Collaborative Sensemaking

Using Knowledge-Transfer Graphs

Knowledge-Transfer Graph (KTGraph) mediates awareness of analysis coverage, supports explicit communication of progress and uncertainty with annotation, and implicit communication through playback of investigation histories.

The paper is organized in the following structure:
Intro
Background and related work
Design consideration
System intro: features, then each panel
Sensemaking tasks (application scenarios)
User study
Discussion
Limitation
Conclusion

VizML: A Machine Learning Approach to Visualization

Recommendation

Authors proposed a machine learning-based approach to visualization recommendation that learns visualization design choices from a large corpus of datasets and associated visualizations. They first extracted features from four categories:

- Dimension: the number of rows in a column.
- Type: categorical, temporal, or quantitative.
- Value: statistical and structural properties of the values within a column.

- Name: the column name.

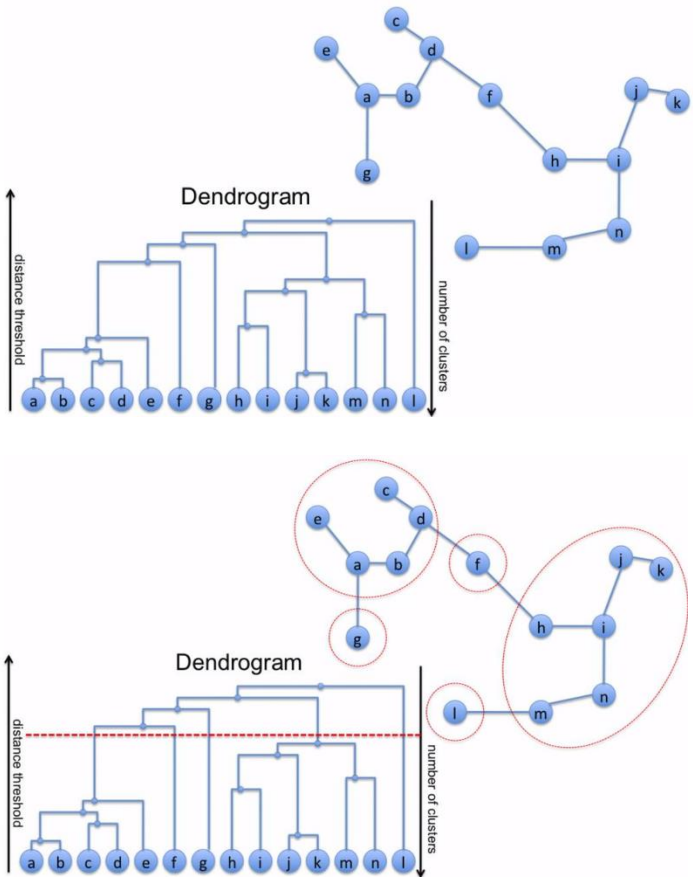
and identified five key design choices made by analysts while creating visualizations:

- Encoding-level design choices: mark types, such as scatter, line, bar; is shared axis and is X- or Y- axis
- Visualization-level design choices of a chart: visualization type; has a single axis

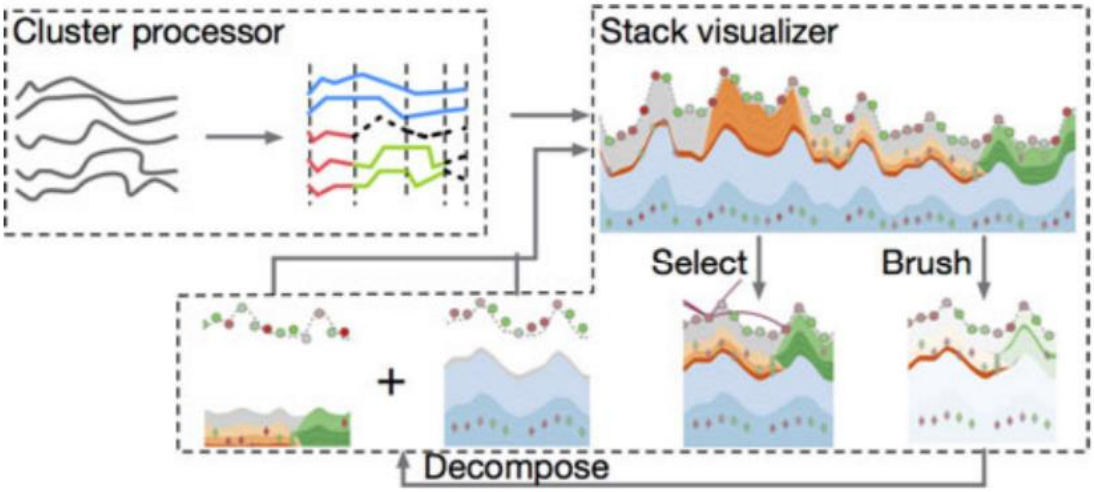
Then, they trained models to predict these design choices using one million dataset-visualization pairs collected from an online visualization platform, called Plotly.

Agglomerative clustering

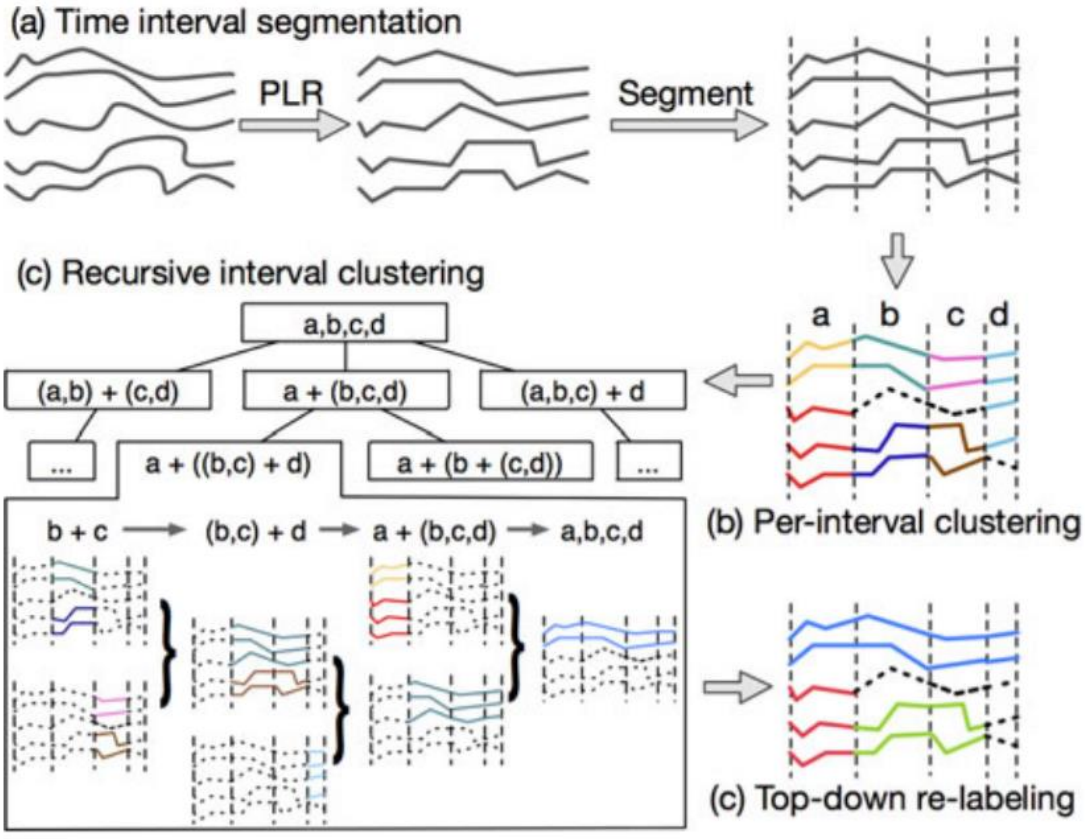
This is a hierarchical clustering approach, which clusters the nodes/clusters according to an increasing distance threshold. The distance between clusters equals to the minimum distance between a pair of nodes from the two clusters. Once the dendrogram is generated, the clustering result can be dynamically adjusted according to the distance threshold or the number of clusters.



PieceStack: Toward Better Understanding of Stacked Graphs



When a collection of temporal sequences is loaded, they are first partitioned and clustered based on their local trend similarities with the clustering processor. The stack visualizer then transforms the output into a comprehensible overview visualization.



The clustering algorithm includes the above steps. The recursive interval clustering recovers the cluster duration information from the interval clustering results. The basic idea is to enumerate all possible ways of grouping the basic intervals, and find the best grouping strategy based on a cost function, which has two considerations:

- Minimizing the number of un-clustered segments.
- Resulting in the smallest possible number of clusters under the above condition.